

The Words and Sounds of Telephone Conversations

By NORMAN R. FRENCH, CHARLES W. CARTER, JR.,
and WALTER KOENIG, JR.

This paper presents data concerning the vocabulary and the relative frequency of occurrence of the speech sounds of telephone conversation. Tables are given showing the most frequently used words, the syllabic structure of the words, the relative occurrences of the sounds, and, for each vowel, the percentage distribution of the consonants which precede and follow it. Comparisons are made with the vocabulary and relative occurrence of speech sounds in written English.

INTRODUCTION

CONVERSATION resembles other forms of communication in its use of symbols, in themselves merely physical phenomena, but which combined in sequence are by convention endowed with meaning. The elementary symbols used in conversation are the acoustic disturbances called speech sounds. A language is characterized by the speech sounds which it uses and by the combinations of speech sounds which form syllables and words. The physical description of a language involves a statement of the characteristics of the individual sounds and also of the frequency of occurrence of each sound and combination of sounds. The latter or statistical aspect of conversation is treated in this paper.¹

Studies of the relative frequency of English speech sounds have been made previously, but they have been confined, so far as the writers have ascertained, to the analysis of written matter. Of these an extended investigation is that made by Godfrey Dewey.² For pedagogical purposes in connection with difficulties in spelling and in developing methods of shorthand writing, which seem to have been the aims in the previous studies, written matter is the natural point of departure.

There are obvious differences between English when read aloud from printed matter and English used as a medium of conversation, which might be expected to produce differences between analyses based on the two forms. Written matter is permanent and, to some degree, self-conscious; it receives qualification by dependent clauses and preposi-

¹ Some of the results of this study were presented at the May, 1929, meeting of the Acoustical Society of America. See French and Koenig, *Journal A. S. of A.*, October, 1929, p. 110.

² "Relative Frequency of English Speech Sounds," *Harvard Studies in Education*, IV. Harvard University Press, 1923.

tional phrases and it makes use of synonyms and a vocabulary more or less ample according to the writer's fancy and ability. In conversation attention seems to be paid more to the thought than the form of expression, with the exception, perhaps, that certain modes acceptable in writing may be considered as too formal for conversation. It is doubtful, however, that conversation should be described as more concise than written matter. The sentences are, indeed, likely to be shorter. They are often incomplete, in fact. But often in conversation even a single statement is completed only after a number of fumbling attempts, an oral manifestation of crystallizing thought, whereas in written matter the final expression alone would appear. In repetition of a thought, synonyms are less likely to be found in conversation than in written matter. Dependent clauses are less frequent than in written matter. Qualification and description often take the form of separate sentences, so that those words characteristic of involved construction tend to be less prominent in conversation, while the framework words, such as the auxiliary verbs and pronouns, are more intensively used. These differences, which tend to restrict the vocabulary, will be found reflected in the comparisons given later in this paper.

The material for the present study was obtained from telephone conversations over typical toll circuits terminating in the city of New York. The process of noting the words of the conversations was carried out in the following manner: During one week the observer recorded nothing but the nouns used, during another week she recorded only verbs, and during a third week only adjectives and adverbs. This routine was repeated until observations had been made on 500 conversations for nouns, 500 conversations for verbs, and 500 conversations for adjectives and adverbs. Three other classes of words were recorded: prepositions and conjunctions, pronouns, and articles; but for these classes approximately 150 conversations in each case were judged to be sufficient.

Certain classes of words were, for various reasons, omitted entirely. These are names, titles, exclamations, letters, numbers and the nameless sound which may be transliterated as "er" or "uh," so frequently punctuating a haltingly expressed sentence. A more comprehensive method, but based on a much smaller number of conversations, indicates that the ratio of the total number of occurrences of words in the omitted classes to the number of occurrences of the words discussed in this paper is about one to four. Within the omitted group the division is roughly as follows: proper names and titles, 20 per cent; exclamations and interjections, such as "yes," "no," "well," "yeah," "uh-huh," "oh," "all right," "hello," "good-by," laughter

and profanity, 40 per cent; letters and numbers, 25 per cent; and the sound "er," 15 per cent.

The words which were obtained by the process of sampling conversations for specific parts of speech are not, of course, identical with those which would have been obtained had the entire conversation been recorded. The representativeness of the most frequent words, which largely determine the relative frequency of the speech sounds, was investigated by a later test in which a different observer recorded the verbs from 250 conversations. These results will be discussed later, but it may be pointed out here that the word list obtained by the two observers corresponded so closely that it is felt that the samples of parts of speech were recorded with sufficient accuracy and were sufficiently large to justify taking the words obtained as a good representation of the main body of telephone conversation.

The kinds of conversations encountered are shown in Table I. The great preponderance of business calls is reflected, as will be shown later,

TABLE I
TYPES OF TELEPHONE CALLS ON WHICH OBSERVATIONS WERE MADE

<i>a. Material</i>	
Business Calls.....	89.0 per cent
All other Calls.....	11.0 per cent
<i>b. Speakers</i>	
Two Men.....	86.5 per cent
Two Women.....	10.4 per cent
Man and Woman.....	3.1 per cent

in the vocabulary. If a smaller percentage of the calls had been business in nature and if a larger percentage had been between women the vocabulary would probably have been different. Whether any marked change would have been found is open to some doubt when it is recalled that business may cover a wide range of topics and that in the 1,900 conversations from which samples were taken there may have been as many as 3,800 different speakers. Evidence will be given, however, which indicates that the relative frequency of the speech sounds would have been changed very little.

WORDS

The number of conversations on which observations were made was regulated to some extent by the ratio of the number of total words to the number of different words recorded in each class. In the early stages of observing many of the total words recorded were different, making this ratio low, but as the observations continued fewer and fewer new words were encountered. In Figure 1 curves are given

which show, for two classes of words, the way in which the number of different words in each class varied as the total number of words in that class increased. To take the nouns, of the first 200 about half were different, of the first 1,000 about a third were different, of the total of 11,660 nouns recorded about one tenth were different. An extrapolation of the curve indicates that the observations would need

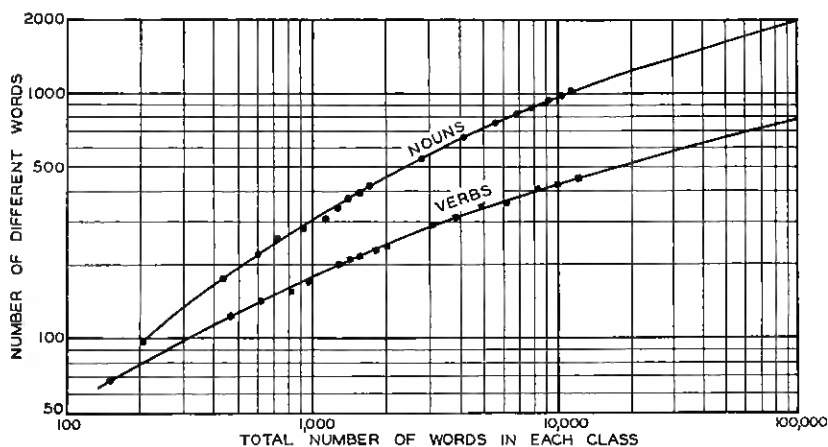


Fig. 1—The number of different words occurring in a given total number of words, for nouns and for verbs.

to be increased tenfold from this point in order to double the number of different nouns. Approximately the same extension of the observations would be required to double the number of different verbs. In neither case, however, could a material change in the relative occurrence of the speech sounds be expected if the observations were so extended. This will be shown below.

Table II shows the total number of words and the number of different words for each part of speech separately. The verbs and auxiliary verbs, which were recorded together, have been separated in the table. The numbers of total words for the other three minor classes have been found by multiplying the observed figures by the ratio of 500 to the actual number of conversations (about 150) on which observations were made for these classes. The numbers of different words for these classes are not similarly increased since virtually all the possible different words were obtained in the observations. In finding the number of different words the various forms of the words, such as the plural form of the nouns, the different tenses of the verbs, and the comparative and superlative forms of the adjectives, have not been counted as separate words, although they were recorded and are

TABLE II
OCCURRENCE OF PARTS OF SPEECH

Parts of Speech	Number of Words		Ratio Total to Different
	Total	Different	
Nouns.....	11,660	1,029	11.3
Adjectives and Adverbs.....	9,880	634	15.6
Verbs.....	12,550	456	27.5
Auxiliary Verbs.....	9,450	37	255.
Pronouns *.....	17,900	45	398.
Prepositions and Conjunctions *.....	12,400	36	344.
Articles *.....	5,550	3	1850.
	79,390	2,240	35.4

* Derived from data on less than 500 conversations.

treated separately in the analysis for speech sounds. An exception to this is that each form of the auxiliary verbs "be," "can," "may," etc., was counted as a separate word.

It is of interest to find that of approximately 80,000 words so obtained, only 2,240, or less than 3 per cent, are different words. If each of the modifications of a word is counted as a different word the number of different words is increased to 2,822; but even on this basis less than 4 per cent of the total words are different words. Even among the nouns the number of different nouns is only a tenth of the total number of nouns. The five minor parts of speech shown in the last four lines of Table II form only 5 per cent of the different words and yet make up 57 per cent of the total words. The nouns, which constitute 46 per cent of the different words, contribute only 15 per cent of the total words. Such figures indicate clearly that conversation is based on a framework built up of a relatively small number of different words, arranged in many patterns, which supports the more variegated words which convey most of the meaning.

A more detailed idea of this framework is given by Tables III-*a* and III-*b*, which contain a list of the words which were observed in at least 1 per cent of the conversations. In Table III-*a* the words are arranged in order according to the total number of times they were recorded. This is approximately, but not quite, the same as the order of the number of conversations in which they occurred as may be seen by examining the numbers following each word. In Table III-*b* the same words are arranged alphabetically, for ease in reference. The list comprises 737 words out of the 2,240 different words recorded. The importance of the list lies in the fact, as will be shown later, that these words almost completely determine the relative frequency with which the elementary

TABLE III-a

WORD LIST—NUMERICAL ORDER *

*Words which Occurred in One Per Cent or More
of the 500 Telephone Conversations Analyzed*

Column A: Total number of times the word (or some form of it) was used.
Column B: Number of conversations in which the word occurred.

A		B	A		B	A		B
3,990	1	467	336	what	193	170	anything	100
3,540	you	499	330	morning	191	170	my	97
3,110	the	496	326	an	178			
2,060	a	487	321	just	211	168	night	107
2,046	on	458	317	over	208	159	call, n.	111
1,942	to	472	296	be	175	157	your	100
1,792	that	397				156	little	117
1,605	it	417	295	or	178	146	stuff, n.	92
1,506	is	419	295	take	207	146	won't	115
1,363	and	391	276	am	172	140	last, a.	106
			274	come	168	140	she	50
1,360	get	393	274	make, v.	169	139	all	100
1,305	will, aux.	402	273	give	172	139	better	103
1,190	of	396	268	very	165			
1,170	in	408	264	send	172	139	number, n.	80
1,115	he	297	262	as	125	138	out	90
1,100	we	294	259	right, a.	173	137	try	100
913	they	253				133	ask	101
887	see	328	247	order, n.	119	133	sell	81
883	have	367	243	good	149	131	not	96
823	for	330	241	minute	155	130	those	100
			241	price	123	125	only	84
753	know	325	238	here	157	121	business	83
640	don't	301	234	car	88	120	office	83
638	do	302	230	had	151			
618	are	293	229	time	165	118	late	94
599	want	297	228	can't	132	118	no, a.	77
597	go	280	226	much	160	117	all right	74
553	tell	264				115	pretty	92
518	with	263	224	there	144	115	shipment	80
496	me	283	222	week	120	113	back, a.	79
486	him	223	215	let	148	112	look, v.	85
			214	letter	112	112	mean, v.	82
480	about	266	209	any	140	112	off	67
476	at	238	200	did	144	109	hear	85
474	think	232	199	more	134			
473	this	240	195	didn't	142	108	ship, v.	68
458	day	251	193	talk, v.	131	108	way	81
418	thing	235	193	today	124	106	his	70
410	say	211				105	dollar	66
396	can, aux.	221	190	other	128	105	too	77
386	call, v.	200	186	company	111	105	wire, n.	78
379	would	207	186	fine, a.	122	104	haven't	83
			184	could	124	104	then	88
370	them	170	183	same	127	103	how	78
358	was	194	179	put	114	103	who	74
339	now	216	178	wait, v.	135			
338	from	196	176	has	114	98	buy	60

* In ambiguous cases the part of speech is denoted as follows: noun, n.; verb, v.; adjective or adverb, a.; auxiliary verb, aux.; preposition, prep.

TABLE III-a (Cont'd)

A		B	A		B	A		B
97	man	67	64	some	43	43	chance	37
97	wouldn't	79	63	been	53	43	coffee	8
96	before	79	63	but	54	43	every	36
96	first	71	63	contract	31	42	stock, n.	33
96	market	51	63	out of	25	42	than	33
93	something	67	63	sample, n.	37			
92	month	70	63	these	57	41	feel	30
92	well, a.	71				40	different	30
89	case	47	61	few, a.	50	40	meet	28
			61	ton	27	40	reason, n.	32
89	find, v.	72	61	train, n.	33	40	show, v.	30
88	by	75	60	best	52	40	which	33
88	probably	69	60	everything	47	40	yesterday	35
87	afternoon	62	60	may	50	39	pound, n.	21
87	line	60	60	thank	56	38	doing	34
87	name, n.	52	59	check, n.	35	38	keep	28
86	like, v.	71	58	along, prep.	33			
86	sure	72	58	job	44	38	old	22
86	yet	67				37	awful	31
85	fellow	70	58	tonight	40	37	bag	24
			58	up, prep.	42	37	certainly	28
85	pay, v.	55	57	home	35	37	difference	31
84	talk, n.	67	57	our	47	37	information	29
84	write	61	56	another	46	37	matter, n.	31
83	new	62	56	away	45	37	must	30
83	next	61	56	should	43	37	phone, n.	35
83	were	66	55	expect	49	37	seem	33
82	understand	63	54	around	54			
82	when	69	54	copy, n.	36	36	boy	31
79	people	59				36	hand, n.	31
78	year	53	54	idea	38	36	hour	30
			53	bad	46	36	house	27
77	us	63	53	couldn't	47	36	mind, n.	30
76	soon	63	52	bill, n.	39	35	early	26
75	place, n.	55	52	nice	38	35	figure, v.	24
73	money	56	52	tomorrow	36	35	oil, n.	18
72	guess, v.	63	52	word	45	35	question, n.	31
71	after	54	51	big	42	35	quite	32
71	hold, v.	60	51	where	42			
71	through	46	51	whole	40	34	ahead	27
69	isn't	52				34	point, n.	27
69	leave	54	50	cent	31	34	wonder	28
			49	figure, n.	32	33	offer, n.	20
68	coal	27	49	glad	39	33	speak	31
68	might, aux.	59	49	ship, n.	28	33	unless	33
68	work, v.	50	48	report, n.	29	32	bid, n.	24
67	again	62	47	suppose	41	32	deliver	24
67	her	30	46	into	42	32	less	23
67	its	67	45	boat	22	32	possible	24
67	so	53	45	couple	38			
67	their	63	45	high	34	31	believe	25
66	long	55				31	check, v.	28
65	because	47	45	ought	37	31	low	22
			45	trouble	33	31	situation	25
65	use, v.	50	44	barrel	20	31	touch, n.	29
65	work, n.	49	44	delivery	36	31	why	25
64	listen	55	43	anybody	40	30	basis	21

TABLE III-a (Cont'd)

A		B	A		B	A		B
30	fix, v.	21	24	hope, v.	19	18	steel	10
30	move	23				18	trip, n.	17
30	ready	25	24	near	23	18	wasn't	18
			24	piece	16	17	above	13
30	receive	22	24	start, v.	22	17	accept	14
30	sorry	25	24	wrong	20	17	against	13
30	town	22	23	busy	17	17	amount	15
29	between	29	23	ever	22	17	appointment	14
29	does	27	23	foot	13	17	cable	10
29	dope, n.	24	23	lot	19	17	cover, v.	14
29	mail, n.	22	22	card	9			
29	many	25	22	forget	18	17	definite	13
29	moment	26				17	goods	13
29	need, v.	22	22	friend	14	17	plant, n.	9
			22	special	15	17	possibility	15
29	paper	17	22	wire, v.	18	17	size	12
29	telegram	19	21	balance, n.	15	17	somebody	17
29	telephone, n.	27	21	change, n.	15	17	still, a.	16
29	though	29	21	loan	5	17	story	12
28	able	26	21	mail, v.	16	17	ticket	9
28	customer	22	21	welcome, a.	21	17	within	17
28	instruction	20	20	account, n.	16			
28	note, n.	24	20	agreement	8	16	handle, v.	14
28	ring, n.	23				16	like, a.	16
28	room	19	20	anyhow	17	16	part, n.	15
			20	cut, v.	18	16	quote	14
28	sale	25	20	exactly	17	16	tank	7
27	arrange	23	20	happen	15	16	truck	13
27	bring	24	20	list, n.	13	15	along, a.	13
27	doesn't	23	20	message	11	15	also	11
27	done	23	20	most	15	15	answer, n.	15
27	maybe	26	20	record, n.	18	15	board	8
27	never	23	20	stop, v.	18			
27	order, v.	24	20	terrible	13	15	cargo	8
27	really	25				15	clean, v.	14
27	share, n.	10	19	address, n.	15	15	clear, a.	14
			19	department	16	15	cocoa	9
27	stay	23	19	far	15	15	cost, v.	15
27	wish, v.	22	19	hold, n.	17	15	date	14
26	book	17	19	load, v.	16	15	interest, n.	9
26	inch	7	19	meeting	9	15	item	10
26	machine	14	19	nearly	19	15	station	8
26	proposition	21	19	plan, n.	12	15	spend	11
26	railroad	19	19	position	15			
26	run, v.	20	19	rate	11	15	worry, v.	14
26	short	21				14	already	14
25	bank	12	19	straight	15	14	arrangement	11
			18	anyway	16	14	bid, v.	11
25	change, v.	22	18	cheap	13	14	club	7
25	city	18	18	even	17	14	extra	11
25	hasn't	25	18	imagine	17	14	fact	14
25	help, v.	19	18	lunch	18	14	finish, v.	10
25	material	14	18	pier	10	14	full	12
24	absolutely	21	18	possibly	14	14	help, n.	10
24	care, v.	24	18	quotation	13			
24	down	20	18	small	17	14	hotel	11
24	hard	22				14	open, a.	12

TABLE III-a (Cont'd)

A		B	A		B	A		B
14	operator	10	12	real	11	10	sheet	8
14	particular	13	12	satisfactory	11	10	street	8
14	perfectly	12	12	several	11	10	territory	5
14	profit	11	12	somewhere	12	10	together	8
14	read	11	12	steamer	10			
14	report, v.	12	12	warehouse	8	10	transfer, n.	8
14	second, n.	12				10	warm	7
14	set, n.	8	11	afraid	11	10	whatever	10
			11	almost	11	10	woman	5
14	sign, v.	12	11	arrive	10	10	yourself	10
14	stand, v.	14	11	both	10	9	build	7
14	surely	14	11	box, n.	7	9	care, n.	6
14	turn, v.	11	11	cold, n.	6	9	careful	9
13	across	13	11	complete, v.	8	9	certain	8
13	answer, v.	9	11	concern, n.	10	9	charge, v.	8
13	bond	8	11	confirm	7			
13	building	11	11	definitely	10	9	color	8
13	charge, n.	8				9	complete, a.	8
13	condition	12	11	detail	10	9	conference	7
			11	drawing	8	9	decide	9
13	connection	12	11	funny	11	9	end, n.	7
13	deal, n.	12	11	light, a.	7	9	express, n.	7
13	direct, a.	11	11	mile	8	9	game	8
13	drop, v.	12	11	motor	7	9	hospital	6
13	further	11	11	personally	8	9	immediately	7
13	general, a.	8	11	quality	10	9	large	8
13	himself	13	11	rather	11			
13	insurance	11	11	use, n.	10	9	mention	7
13	interested	10				9	necessary	9
13	least	12	10	air	6	9	outside	9
			10	awfully	10	9	personal	9
13	luck	12	10	bother, v.	9	9	remember	8
13	notify	6	10	carload	9	9	sit	8
13	offer, v.	12	10	cold, a.	7	9	sometime	9
13	party	12	10	crazy	8	9	statement	9
13	person	13	10	dinner	7	9	suggestion	8
13	quick	13	10	double	7	9	supply, v.	7
13	test, n.	8	10	easily	9			
13	without	13	10	either	9	9	true	9
12	agree	11				9	up, a.	8
12	always	10	10	enough	10	9	weren't	7
			10	everybody	10	9	willing	7
12	appreciate	11	10	explain	9	9	wise	7
12	bed	10	10	final	6	8	additional	8
12	brother	11	10	freight	8	8	advise	7
12	close, v.	11	10	having	10	8	agent	6
12	consider	9	10	head	9	8	agreeable	7
12	else	12	10	important	10	8	anxious	7
12	expense	10	10	kind, n.	9			
12	fair	12	10	limit, n.	8	8	average, n.	7
12	great	11				8	beyond	8
12	loss	10	10	load, n.	8	8	carry	7
			10	mark, n.	8	8	certificate	5
12	original	10	10	particularly	9	8	close, a.	8
12	per cent	8	10	positively	10	8	each	8
12	pick, v.	11	10	power	5	8	easy	7
12	policy	6	10	service	10	8	engineer	5

TABLE III-a (Cont'd)

A		B	A		B	A		B
8	except	8	7	locate	7	6	offhand	6
8	fill	8	7	lovely	6	6	operate	6
			7	mind, v.	7	6	opportunity	6
8	firm, a.	5	7	mother	7	6	package	6
8	girl	6	7	once	5	6	practically	6
8	guarantee, n.	7	7	ours	7	6	promise	6
8	heavy	6				6	realize	5
8	look, n.	8	7	phone, v.	6	6	represent	6
8	middle, a.	7	7	proper	7	6	shall	6
8	mistake, n.	7	7	sake	6	6	simple	6
8	news	7	7	satisfied	7			
8	ordinary	6	7	side	7	6	straighten	6
8	owe	6	7	state, n.	6	6	such	6
			7	store, n.	5	6	thanks	6
8	plan, v.	8	7	supply, n.	7	6	touch, v.	6
8	push, v.	6	7	throat	5	6	unload	5
8	quantity	6	7	wonderful	6	5	advisable	5
8	reasonable	7				5	allow	5
8	regular	8	7	yard	5	5	approval	5
8	reply, n.	7	6	advice	6	5	catch	5
8	sail, v.	7	6	afford	5	5	conversation	5
8	second, a.	7	6	appear	5			
8	settle	7	6	argument	6	5	correct	5
8	shape	8	6	begin	6	5	crowd	5
			6	broker	5	5	difficulty	5
8	simply	8	6	bunch	5	5	disappointed	5
8	single	7	6	cancel	5	5	discuss	5
8	suggest	8	6	claim, v.	5	5	doctor	5
8	sweet	7				5	estimate, v.	5
8	weather	5	6	clear, v.	5	5	grade	5
8	weight	5	6	collect	6	5	holiday	5
8	whether	7	6	competition	5	5	increase, v.	5
8	world	8	6	cost, n.	5			
7	actual	5	6	dandy, a.	6	5	inform	5
7	ago	5	6	dealer	5	5	insist	5
			6	delay, v.	6	5	instead	5
7	apparently	6	6	depend	6	5	intend	5
7	available	5	6	fairly	6	5	interesting	5
7	buyer	5	6	form, n.	5	5	mix	5
7	clean, a.	7				5	operation	5
7	cover, n.	7	6	impossible	5	5	pardon, n.	5
7	desk	7	6	indeed	6	5	payment	5
7	evening	7	6	inquiry	6	5	reach	5
7	event	7	6	issue, n.	5			
7	evidently	7	6	lay	6	5	reduction	5
7	exact	7	6	lose	5	5	return	5
			6	mark, v.	6	5	show, n.	5
7	favor	7	6	memorandum	6	5	sort, n.	5
7	follow	7	6	notice, n.	6	5	specification	5
7	indicate	6	6	notice, v.	6	5	surprised	5
7	life	7				5	until	5

TABLE III-b

WORD LIST--ALPHABETICAL ORDER *

*Words Which Occurred in One Per Cent or More
of the 500 Telephone Conversations Analyzed*

Column A: Total number of times the word (or some form of it) was used.
Column B: Number of conversations in which the word occurred.

A		B	A		B	A		B
	A		18	anyway	16	10	bother, v.	9
			7	apparently	6	11	box, n.	7
2,060	a	487	6	appear	5	36	boy	31
28	able	26	17	appointment	14	27	bring	24
480	about	266	12	appreciate	11	6	broker	5
17	above	13	5	approval	5	12	brother	11
24	absolutely	21	618	are	293	9	build	7
17	accept	14	6	argument	6	13	building	11
20	account, n.	16	54	around	54	6	bunch	5
13	across	13	27	arrange	23	121	business	83
7	actual	5	14	arrangement	11	23	busy	17
8	additional	8	11	arrive	10	63	but	54
19	address, n.	15	262	as	125	98	buy	60
6	advice	6	133	ask	101	7	buyer	5
5	advisable	5	476	at	238	88	by	75
8	advise	7	7	available	5			
6	afford	5	8	average, n.	7		C	
11	afraid	11	56	away	45			
71	after	54	37	awful	31	17	cable	10
87	afternoon	62	10	awfully	10	386	call, v.	200
67	again	62				159	call, n.	111
17	against	13		B		396	can, aux.	221
8	agent	6				6	cancel	5
7	ago	5	113	back, a.	79	228	can't	132
12	agree	11	53	bad	46	234	car	88
8	agreeable	7	37	bag	24	22	card	9
20	agreement	8	21	balance, n.	15	24	care, v.	24
34	ahead	27	25	bank	12	9	care, n.	6
10	air	6	44	barrel	20	9	careful	9
139	all	100	30	basis	21	15	cargo	8
5	allow	5	296	be	175	10	carload	9
117	all right	74	65	because	47	8	carry	7
11	almost	11	12	bed	10	89	case	47
58	along, prep.	33	63	been	53	5	catch	5
15	along, a.	13	96	before	79	50	cent	31
14	already	14	6	begin	6	9	certain	8
15	also	11	31	believe	25	37	certainly	28
12	always	10	60	best	52	8	certificate	5
276	am	172	139	better	103	43	chance	37
17	amount	15	29	between	29	25	change, v.	22
326	an	178	8	beyond	8	21	change, n.	15
1,363	and	391	32	bid, n.	24	13	charge, n.	8
56	another	46	14	bid, v.	11	9	charge, v.	8
15	answer, n.	15	51	big	42	18	cheap	13
13	answer, v.	9	52	bill, n.	39	59	check, n.	35
8	anxious	7	15	board	8	31	check, v.	28
209	any	140	45	boat	22	25	city	18
43	anybody	40	13	bond	8	6	claim, v.	5
20	anyhow	17	26	book	17	15	clean, v.	14
170	anything	100	11	both	10	7	clean, a.	7

* In ambiguous cases the part of speech is denoted as follows: noun, n.; verb, v.; adjective or adverb, a.; auxiliary verb, aux.; preposition, prep.

TABLE III-b (Cont'd)

A		B	A		B	A		B
15	clear, a.	14	195	didn't	142	7	favor	7
6	clear, v.	5	37	difference	31	41	feel	30
12	close, v.	11	40	different	30	23	foot	13
8	close, a.	8	5	difficulty	5	85	fellow	70
14	club	7	10	dinner	7	61	few, a.	50
68	coal	27	13	direct, a.	11	49	figure, n.	32
15	cocoa	9	5	disappointed	5	35	figure, v.	24
43	coffee	8	5	discuss	5	8	fill	8
10	cold, a.	7	638	do	302	10	final	6
11	cold, n.	6	5	doctor	5	89	find, v.	72
6	collect	6	29	does	27	186	fine, a.	122
9	color	8	27	doesn't	23	14	finish, v.	10
274	come	168	38	doing	34	8	firm, a.	5
186	company	111	105	dollar	66	96	first	71
11	complete, v.	8	27	done	23	30	fix, v.	21
9	complete, a.	8	640	don't	301	7	follow	7
6	competition	5	29	dope, n.	24	823	for	330
11	concern, n.	10	10	double	7	22	forget	18
13	condition	12	24	down	20	6	form, n.	5
9	conference	7	11	drawing	8	10	freight	8
11	confirm	7	13	drop, v.	12	22	friend	14
13	connection	12				338	from	196
12	consider	9		E		14	full	12
63	contract	31				11	funny	11
5	conversation	5	8	each	8	13	further	11
54	copy, n.	36	35	early	26			
5	correct	5	10	easily	9		G	
15	cost, v.	15	8	easy	7			
6	cost, n.	5	10	either	9	9	game	8
184	could	124	12	else	12	13	general, a.	8
53	couldn't	47	9	end, n.	7	1,360	get	393
45	couple	38	8	engineer	5	273	give	172
7	cover, n.	7	10	enough	10	8	girl	6
17	cover, v.	14	5	estimate, v.	5	49	glad	39
10	crazy	8	18	even	17	597	go	280
5	crowd	5	7	evening	7	243	good	149
28	customer	22	7	event	7	17	goods	13
20	cut, v.	18	23	ever	22	5	grade	5
			43	every	36	12	great	11
	D		10	everybody	10	8	guarantee, n.	7
			60	everything	47	72	guess, v.	63
			7	evidently	7			
6	dandy, a.	6	7	exact	7		H	
15	date	14	7	exactly	17			
458	day	251	20	except	8	230	had	151
13	deal, n.	12	8	expect	49	36	hand, n.	31
6	dealer	5	55	expense	10	16	handle, v.	14
9	decide	9	12	explain	9	20	happen	15
17	definite	13	10	express, n.	7	24	hard	22
11	definitely	10	9	extra	11	176	has	114
6	delay, v.	6	14			25	hasn't	25
32	deliver	24		F		883	have	367
44	delivery	36				104	haven't	83
19	department	16				10	having	10
6	depend	6	14	fact	14	10	head	9
7	desk	7	12	fair	12	10	hear	85
11	detail	10	6	fairly	6	109	heavy	6
200	did	144	19	far	15	8		

TABLE III-b (Cont'd)

A		B	A		B	A		B
1,115	he	297		K		112	mean, v.	82
25	help, v.	19				40	meet	28
14	help, n.	10	38	keep	28	19	meeting	9
67	her	30	10	kind	9	6	memorandum	6
238	here	157	753	know	325	9	mention	7
45	high	34				20	message	11
486	him	223		L		8	middle, a.	7
13	himself	13				68	might, aux.	59
106	his	70	9	large	8	11	mile	8
71	hold, v.	60	140	last, a.	106	36	mind, n.	30
19	hold, n.	17	118	late	94	7	mind, v.	7
5	holiday	5	6	lay	6	241	minute	155
57	home	35	13	least	12	8	mistake, n.	7
24	hope, v.	19	69	leave	54	5	mix	5
9	hospital	6	32	less	23	29	moment	26
14	hotel	11	215	let	148	73	money	56
36	hour	30	214	letter	112	92	month	70
36	house	27	7	life	7	199	more	134
103	how	78	11	light, a.	7	330	morning	191
			86	like, v.	71	20	most	15
	I		16	like, a.	16	7	mother	7
			10	limit, n.	8	11	motor	7
3,990	I	467	87	line, n.	60	30	move	23
54	idea	38	20	list, n.	13	226	much	160
18	imagine	17	64	listen	55	37	must	30
9	immediately	7	156	little	117	170	my	97
10	important	10	10	load, n.	8		N	
6	impossible	5	19	load, v.	16			
1,170	in	408	21	loan	5			
26	inch	7	7	locate	7	87	name, n.	52
5	increase, v.	5	66	long	55	24	near	23
6	indeed	6	112	look, v.	85	19	nearly	19
7	indicate	6	8	look, n.	8	9	necessary	9
5	inform	5	6	lose	5	29	need, v.	22
37	information	29	12	loss	10	27	never	23
6	inquiry	6	23	lot	19	83	new	62
5	insist	5	7	lovely	6	8	news	7
5	instead	5	31	low	22	83	next	61
28	instruction	20	13	luck	12	52	nice	38
13	insurance	11	18	lunch	18	168	night	107
5	intend	5				118	no, a.	77
15	interest, n.	9		M		131	not	96
13	interested, a.	10	26	machine	14	28	note, n.	24
5	interesting, a.	5	29	mail, n.	22	6	notice, v.	6
46	into	42	21	mail, v.	16	6	notice, n.	6
1,506	is	419	274	make, v.	169	339	notify	6
69	isn't	52	97	man	67	139	now	216
6	issue, n.	5	29	many	25		number, n.	80
1,605	it	417	10	mark, n.	8		O	
15	item	10	6	mark, v.	6			
67	its	67	96	market	51	1,190	of	396
			25	material	14	112	off	67
	J		37	matter, n.	31	33	offer, n.	20
			60	may	50	13	offer, v.	12
58	job	44	27	maybe	26	6	offhand	6
321	just	211	496	me	283	120	office	83

TABLE III-b (Cont'd)

A		B	A		B	A		B
17	still, a.	16	71	through	46	222	week	120
42	stock, n.	33	17	ticket	9	8	weight	5
20	stop, v.	18	229	time	165	21	welcome, a.	21
7	store, n.	5	1,942	to	472	92	well, a.	71
17	story	12	193	today	124	83	were	66
19	straight	15	10	together	8	9	weren't	7
6	straighten	6	52	tomorrow	36	336	what	193
10	street	8	60	ton	27	10	whatever	10
146	stuff, n.	92	58	to-night	40	82	when	69
6	such	6	105	too	77	51	where	42
8	suggest	8	31	touch, n.	29	9	whether	7
9	suggestion	8	6	touch, v.	6	40	which	33
9	supply, v.	7	30	town	22	103	who	74
7	supply, n.	7	61	train, n.	33	51	whole	40
47	suppose	41	10	transfer, n.	8	31	why	25
86	sure	72	18	trip, n.	17	1,305	will, aux.	402
14	surely	14	45	trouble	33	9	willing	7
5	surprised	5	16	truck	13	105	wire, n.	78
8	sweet	7	9	true	9	22	wire, v.	18
			137	try	100	9	wise	7
	<i>T</i>		14	turn, v.	11	27	wish, v.	22
						518	with	263
295	take	207		<i>U</i>		17	within	17
193	talk, v.	131				13	without	13
84	talk, n.	67	82	understand	63	10	woman	5
16	tank	7	33	unless	33	7	wonder	28
29	telegram	19	6	unload	5	34	wonderful	6
29	telephone, n.	27	5	until	5	146	won't	115
553	tell	264	58	up, prep.	42	52	word	45
20	terrible	13	9	up, a.	8	68	work, v.	50
10	territory	15	77	us	63	65	work, n.	49
13	test, n.	8	11	use, n.	10	8	world	8
42	than	33	65	use, v.	50	15	worry, v.	14
60	thank	56				379	would	207
6	thanks	6		<i>V</i>		97	wouldn't	79
1,792	that	397				84	wouldn't	61
3,110	the	496	268	very	165	24	wrong	20
67	their	63						
370	them	170		<i>W</i>		<i>Y</i>		
104	then	88						
224	there	144	178	wait, v.	135	7	yard	5
63	these	57	599	want	297	78	year	53
913	they	253	12	warehouse	8	40	yesterday	35
418	thing	235	10	warm	7	86	yet	67
474	think	232	358	was	194	3,540	you	499
473	this	240	18	wasn't	18	157	your	100
130	those	100	108	way	81	10	yourself	10
29	though	29	1,100	we	294			
7	throat	5	8	weather	5			

speech sounds occur. They form 96 per cent of the total occurrences of the words. It is to be noticed that no word was observed to occur in all the conversations.

Of the 1,503 different words not shown on the list, 819 were observed only once and 320 only twice. It is quite likely that if the

observations were repeated this part of the list would be duplicated very imperfectly, since these words, while in general well-known, tend to be technical or specific, hence dependent on particular types of subject matter. All except ten of the omitted words are nouns, verbs, adjectives or adverbs.

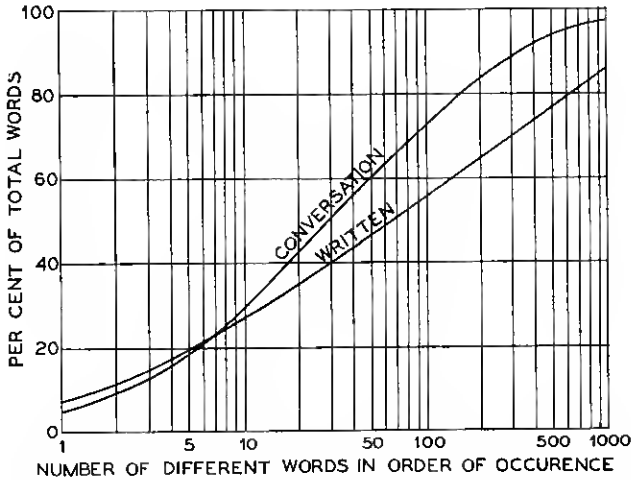


Fig. 2—The cumulative curve obtained when the different words are arranged in order of occurrence.

The importance of a relatively small number of different words which are used very frequently is shown graphically in Fig. 2. The curves shown are cumulative, giving the percentage of the total words contributed by the different words when arranged in the order of their occurrence. The curve labeled "Written" is based on the list given in the study by Dewey, cited above. The economy exercised in conversation, or the poverty of conversational expression, according to the point of view, contrasts sharply with written English. In conversation 30 words account for half the total, in written English 69 words; in conversation 155 words form 80 per cent of the total, in written English 640.

The 50 most common words in telephone conversation and in written English are shown in Table IV, arranged in their order of frequency of occurrence. These words form 60 per cent of the total in conversation and 46 per cent in written English. There are 29 words which are common to the two lists. The personal nature of telephone conversation is shown in the two words which head the list. The most striking difference between the two is the large number of active verbs which occur in the list for conversation: "get," "see," "know," etc.,

TABLE IV
FIFTY COMMONEST WORDS IN TELEPHONE CONVERSATION
Compared with Written English

	Telephone Conversation	Written English		Telephone Conversation	Written English
1.	I	the	26.	GO	HIS
2.	you	of	27.	TELL	BUT
3.	the	and	28.	with	they
4.	a	to	29.	me	ALL
5.	on	a	30.	HIM	OR
6.	to	in	31.	ABOUT	WHICH
7.	that	that	32.	at	will
8.	it	it	33.	THINK	from
9.	is	is	34.	this	HAD
10.	and	I	35.	DAY	HAS
11.	GET	for	36.	THING	ONE
12.	will	be	37.	SAY	OUR
13.	of	was	38.	CAN	an
14.	in	AS	39.	CALL	BEEN
15.	he	you	40.	would	NO
16.	we	with	41.	THEM	THEIR
17.	they	he	42.	was	THERE
18.	SEE	on	43.	NOW	WERE
19.	have	have	44.	from	SO
20.	for	BY	45.	what	MY
21.	KNOW	NOT	46.	MORNING	IF
22.	DON'T	at	47.	an	me
23.	DO	this	48.	JUST	what
24.	are	are	49.	OVER	would
25.	WANT	we	50.	be	WHO

The 21 words not common to both lists appear in capital letters.

12 in all. None of these appears among the 50 commonest words of written English. Three nouns, "day," "thing" and "morning," appear in the conversational list, none in the other. Only one conjunction is found in the conversational list, while five appear in the list for written English.

When the first 100 words in telephone conversation are compared with the first 100 in written English two somewhat unexpected facts emerge. In telephone conversation 14 out of the first 100 are words of more than one syllable; in written English there are ten. Four two-syllable words appear among the first 50 telephone words; the first 59 of written English are monosyllables. A more striking difference concerns the origin of the words. Among the first 100 telephone words there are 11 which are derived through old French from the Latin; in written English there are only two from the Latin. Six of the 11 words occur in the first 65 telephone words, while the first word of Latin origin in written English is the 70th. The telephone words of Latin origin are, in order of occurrence: "just," "very," "order," "minute,"

"price," "car," "letter," "fine," "company," "stuff," "number"; in written English these words are "people" and "very." The predominance of business words in this list for telephone conversation suggests the influence of trade between England and France in the Middle Ages.

More detailed comparisons may be drawn from Table V, which lists the first 25 nouns, the first 25 verbs and the first 25 adjectives and

TABLE V
TWENTY-FIVE COMMONEST WORDS BY PARTS OF SPEECH
Compared with Written English

Nouns		Verbs		Adjectives and Adverbs	
Telephone Conversation	Written English	Telephone Conversation	Written English	Telephone Conversation	Written English
day	man	get	say	now	not
thing	time	see	make	just	all
MORNING	WAR	know	come	very	no
ORDER	PEOPLE	want	take	RIGHT	SO
MINUTE	day	go	know	good	WHEN
PRICE	YEAR	tell	go	HERE	any
CAR	thing	think	see	MUCH	more
time	way	say	get	THERE	now
WEEK	WORLD	call	give	any	UP
LETTER	COUNTRY	take	think	more	out
COMPANY	PART	make	LIKE	TODAY	other
NIGHT	business	come	tell	other	only
CALL	LIFE	give	USE	FINE	GREAT
STUFF	FACT	SEND	call	SAME	SOME
NUMBER	LINE	LET	want	little	HOW
business	GUN	TALK	GOVERN	LAST	very
OFFICE	case	PUT	STAND	BETTER	SUCH
SHIPMENT	HOME	WAIT	ask	all	FIRST
way	CENT	TRY	SEEM	out	good
WIRE	POWER	ask	SHOW	not	EVERY
DOLLAR	PRESENT	SELL	look	only	THEN
man	HOUSE	look	NEED	LATE	little
MARKET	LOSS	MEAN	SAVE	no	here
month	month	HEAR	WORK	ALL RIGHT	just
case	PEACE	SHIP	BELIEVE	PRETTY	WELL

The words not common to both lists appear in capital letters.

adverbs, for both telephone conversation and written matter. Among the nouns only eight are common to the two lists. The effects of business are apparent in the telephone list. On the other hand, the nouns of the written English list reflect the fact, pointed out by Dewey, that the list was obtained from a study made soon after the war. Among the verbs 15 words are common to the two lists and those which differ are concentrated at the end. Approximately half the adjectives and adverbs appear in both lists. The nouns from telephone conver-

sation shown in this table form 2.4 per cent of the different nouns and 40 per cent of the total nouns; the verbs form 5.5 per cent of the different verbs and 72 per cent of the total verbs; while the adjectives and adverbs form 3.9 per cent of the different adjectives and adverbs, but 48 per cent of the total.

An examination of the origin of the words in Table V shows that the influence of Latin on the frequently used words is largely confined to nouns. Eleven of the first 25 nouns of telephone conversation, and eight of the first 25 nouns of written English come from the Latin. Among the first 25 telephone nouns, aside from the eight nouns mentioned above among the first 100 words, there are: "office," "market" and "case"; among the first 25 nouns of written English the following are of Latin origin: "people," "country," "part," "fact," "cent," "power," "present" and "peace." Only one of the first 25 telephone verbs comes from Latin: "try," and three of those in written English: "use," "govern" and "save." Among the adjectives and adverbs there are found in the telephone list: "just," "very" and "fine," as above, and in the written English list the word "just" is added to "very," which was in the first 100 words.

Referring once more to the small number of different words found it may be pointed out that this shows how difficult it would be to estimate the size of vocabularies by recording spoken words. The 80,000 words of this study are equivalent to a complete record of seven hours' conversation, taking a rate of 200 words per minute. As noted before, the number of different words was only 2,240, even though the conversations covered a wide range of topics by many different speakers. To increase this number notably, the curves of Figure 1 indicate that the observations would need to be very extended, since the rate at which new words appear has already become very low. For example, if the conversations were to go on continuously for a week at the above rate a total of 2,000,000 words might be expected. By extrapolating the curves of Fig. 1, and using a similar curve for adjectives and adverbs, which lies between the curves shown, it may be estimated that only about 5,000 of these words would be different words. Extrapolation is a rough tool, but even with its inaccuracies in mind, the conclusion seems safe that to measure a vocabulary by recording spoken words involves the risk of gross underestimation unless the observations are exceedingly prolonged.

It is suggested that teachers of languages may find the 737 words in Tables III-*a* and III-*b* to be of practical use in their profession. Presumably the progress of a student in speaking a foreign language would be materially assisted by a thorough knowledge, early in his course,

of the words which are met with great frequency. The present methods of teaching the spoken language no doubt approximate to this, as a result of experience. It is suggested that the present word list, which contains the words used so frequently as to form 96 per cent of the total number observed in this study, provides a guide for the selection of important words to be taught. Additions are needed to the list as it stands, in order to care for certain obvious situations not encountered in telephone conversation concerning, for example, hotels, restaurants and trains. With these points in mind, the list given has the advantage of being founded on a study of actual conversation.

SYLLABLES

As a preliminary to analysis of the words into their component sounds the words were divided into syllables. With regard to the fact that the study concerned conversation the division was made on phonetic lines, which, as unabridged dictionaries show, differ from the orthographical divisions. Likewise a few words such as "every," "preference," "average" and the like were divided into two syllables, according to the usual colloquial pronunciation.

TABLE VI
THE SYLLABIC STRUCTURE OF CONVERSATIONAL VOCABULARY

Parts of Speech	Per Cent of Words Having Number of Syllables Shown						Average Number of Syllables
	1	2	3	4	5	6	
Nouns.....	53.3	33.8	9.7	2.7	0.47	0.03	1.63
Verbs.....	81.9	15.0	2.8	0.3	—	—	1.21
Adjectives and Adverbs.....	57.8	30.7	8.0	2.8	0.66	0.02	1.58
Minor.....	94.8	4.7	0.6	0.1	—	—	1.06
All Words.....	82.0	13.8	3.2	0.86	0.15	0.01	1.23

In Table VI a summary is given of the syllabic structure of words, based on the total occurrence of the words. It may be noticed that words longer than two syllables make up only a trifle more than 4 per cent of the words observed. Nouns tend to be more polysyllabic than other classes, but even so the nouns having more than two syllables occur so infrequently as to form only 13 per cent of all the nouns observed.

The types of phonetic syllables which are found range in complexity from a single vowel through various combinations of consonants with a vowel. The relative number of the different types is shown in Table VII. The letters V and C represent "vowel" and "consonant,"

TABLE VII
TYPES OF PHONETIC SYLLABLES IN TELEPHONE CONVERSATION
Relative Occurrence per Hundred

Type	Occurrence
V	9.7
VC	20.3
CV	21.8
CVC	33.5
VCC	2.8
CCV	0.8
CVCC	7.8
CCVC	2.8
CCVCC	00.5
	<hr/> 100.0

respectively, and the letters CC are used to denote a compound consonant form, that is, two or more consecutive consonants. It may be seen that the typical syllable is of the CVC type, closely followed in importance by the CV and VC types. The syllables having two or more consecutive consonants form about one seventh of the total.

SPEECH SOUNDS

The analysis of the words into their constituent sounds was attended by certain difficulties which should be borne in mind in considering the tables which follow. It was not feasible to record the original words phonetically, just as they were pronounced by the telephone subscriber. Instead the words were recorded and their phonetic values assigned later. In so doing the dictionary was not adopted as an authority for the pronunciation since in the informality of conversation, even among educated persons, there are elisions and changes of stress which cause departures from the dictionary standard. Certain very common words, for example, receive various treatments in conversation, depending on their situation in the sentence, the emphasis desired and the speed of talking. The word "and" may be pronounced as spelled, but quite often it is reduced to " 'nd" or even " 'n'." The prepositions "to" and "of" are similarly varied. Altogether about 40 common words were found, of this type, each of which seemed subject to several different pronunciations, even in speech which would not be regarded as unduly careless. These were all from the minor classes: auxiliary verbs, pronouns, prepositions and conjunctions. The modification, in general, is such as to give the vowel its unstressed value. In the analysis these different forms are included, the weighting for each modification necessarily being a matter of judgment. The remaining words were each assigned a single pronunciation, selecting that which we regarded as being the typical pronunciation

heard in reasonably enunciated conversation among educated persons in New York. The departures from dictionary standards are largely confined to the vowels. As a result the analysis is affected to some degree by the speech habits of the writers.³ It is regrettable that some arbitrariness should be introduced, but this seems to be a difficulty common to discussions of vowel sounds. Some of the difficulty is avoided by making separate classifications for vowels for which the pronunciation is indefinite, such as the vowels in unstressed positions. The articles "the," "a" and "an" were omitted entirely from the analysis on account of the large number of variant pronunciations to which they are subject.

The results of the analysis into speech sounds are shown in Table VIII. Three divisions are given: vowels, initial consonants and final consonants, based on the division into phonetic syllables. The method followed was: first, to divide the words into phonetic syllables, second, to assign phonetic symbols to the sounds and third, to weight each sound by the total number of times the word was recorded. The sounds are identified in the table, where necessary, by key words.

No difficulties were encountered in analysis of the consonants, but a few special points which arose in assigning the vowel qualities may be noted. The key word "pot" is used to denote a vowel sound which is pronounced differently by many natives of New England and those whose habits of speech were formed elsewhere.⁴ With these New Englanders the sound tends toward the quality of the vowel in "pawm," although shorter in duration. The same New Englanders make a real distinction between the vowel of "pot" and the vowel of "palm." By many speakers elsewhere no such distinction is made and the two are lumped into a single intermediate sound which is neither the New Englander's "pot" nor "palm." To avoid confusion the class denoted by "pot" has been made to include "not" and many other monosyllables of the same ending, as well as "on," "job," "stock," etc., which grouping is believed to be homogeneous on either basis. The few words of the class of "palm" which were encountered have been included under "par." The class denoted by "par" may be subdivided into: "par," 1.24; "palm," 0.07. The class denoted by "palm" would be somewhat larger if the class which we may denote by "path," such as "can't," "last," "ask," etc., had not been classified under

³ For the benefit of phoneticians who may be interested it may be stated that the writers are residents of Greater New York of more than six years' standing, that their boyhoods were spent in Maine, Illinois and New Jersey, respectively, and their college years at Maine and Princeton, Harvard and Oxford, New York University and Harvard, respectively; this seems a background sufficiently varied to bring to light many of the principal variants of American speech.

⁴ Just what the geographical lines may be, the writers do not pretend to know. A phonetic map would be of interest.

TABLE VIII

RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATION

All Words (Except Articles)

<i>Vowels</i>		<i>Initial Consonants</i>		<i>Final Consonants</i>	
pin.....	10.27	W	9.38	t	14.30
pine.....	7.58	T	7.86	r	13.05
pan.....	6.89	TH'' (then).....	6.72	n	12.52
pen.....	6.60	Y	6.48	l	8.40
peel.....	6.44	D	6.21	z	6.01
pool.....	6.26	M	5.89	m	5.48
pot.....	5.21	H	5.75	d	4.44
pane.....	4.78	K	5.55	v	4.23
pole.....	4.74	S	5.46	ng	3.57
pawn.....	4.15	N	4.99	s	3.13
pun.....	4.14	B	4.64	k	2.85
pull.....	2.96	G(gun).....	4.33	f	1.37
pout.....	1.69	L	4.31	th'' (with).....	1.25
par.....	1.31	F	3.96	p	1.24
pair.....	1.09	R	2.78	ch53
purr.....	.80	P	2.54	b42
pew.....	.26	TH' (thin).....	2.02	g38
poise.....	.19	SH	1.74	sh32
		V	1.25	j14
	75.36	J83	th'(myth).....	.04
		CH55	zh (azure).....	.01
<i>Unaccented Vowels</i>		Z34	h	—
possible.....	5.52	ZH02	w	—
about.....	5.33	NG	—	y	—
differ.....	4.56				
receive.....	3.78		93.60		83.68
notion.....	2.65	<i>Compounds</i>		<i>Compounds</i>	
wanted.....	1.83	PR	1.06	nt	4.40
people.....	.97	HW91	nd	2.56
	24.64	ST87	st	1.18
		TR69	ts	1.11
		FR62	nk76
		PL36	ld75
	100.00	KW28	rz57
Total Number of		BL23	ks47
Sounds.....	92,522	SP19	kt42
		KL18	rd37
		Others	1.01	Others	3.73
			6.40		16.32
			100.00		100.00
			64,043		65,544

"pan," such being the more common American pronunciation. Actually the occurrence of words in the class of "path" is not high; if they had been given a special class in Table VIII their relative occurrence figure would have been 0.78, reducing the figure for "pan" to 6.11. Special categories are given to the vowel sounds in the classes denoted by "pair" and "purr" since there is often disagreement con-

cerning the quality of a vowel which precedes "r." Likewise it was found expedient to make a number of classifications of vowels in unaccented positions.

Since the figures of Table VIII are likely to find application as weighting factors it is convenient to have them add exactly to 100 per cent, consequently they are given to two places of decimals. An estimate of the representativeness of these figures may be obtained from the data presented in Table IX, which were worked out from

TABLE IX
COMPARISON WITH CHECK TEST
Relative Occurrence of Consonants in Verbs

Sound	First Observations	Check Test	Difference	Sound	First Observations	Check Test	Difference
B	1.02	1.02	.00	S	10.31	9.67	-.64
D	4.46	4.83	+.37	T	16.97	17.39	+.42
F	1.73	2.18	+.45	V	2.36	2.20	-.16
G	11.15	9.40	-1.75	W	4.87	4.54	-.33
H	1.40	1.66	+.26	Y	.55	.53	-.02
J	.22	.23	+.01	Z	1.09	1.44	+.35
K	8.90	8.74	-.16	CH	.32	.75	+.43
L	7.70	7.94	+.24	SH	1.35	1.17	-.18
M	4.45	3.96	-.49	TH'	2.53	2.85	+.32
N	6.87	7.10	+.23	TH''	.05	.06	+.01
P	3.34	3.47	+.13	ZH	.00	.00	.00
R	3.97	3.88	-.09	NG	4.39	5.00	+.61
					100.00	100.00	

observations mentioned before, conducted by a different observer at a different time, but on the same set of toll circuits. Records were made only of verbs, and for 250 instead of 500 conversations. The vocabulary collected in the check test resembled that of the first observations closely. Arranging the words in the order of occurrence, the first 17 words of the first observations are also the first 17 of the check test, although the order is not repeated exactly. In the first observations the first few words run: "get," "know," "see," "want," "go," "tell," "think" and "say"; in the check test the order is: "get," "see," "know," "want," "tell," "think," "go" and "say." Table IX shows the analysis of the words as to the simple consonants, lumping initial and final consonants together. Only one of the differences is greater than 1 per cent and all but three are less than 0.5 per cent. One check test is not sufficient for a final statement, but judging by these results the observing method and the samples taken seem to justify considering the figures of Table VIII as representative as far as the figures in the

digits position for most of the sounds and as to order of magnitude for the infrequent sounds.

The effects of restricting the word list in various ways are shown in Figure 3. The first line shows graphically the relative occurrence per

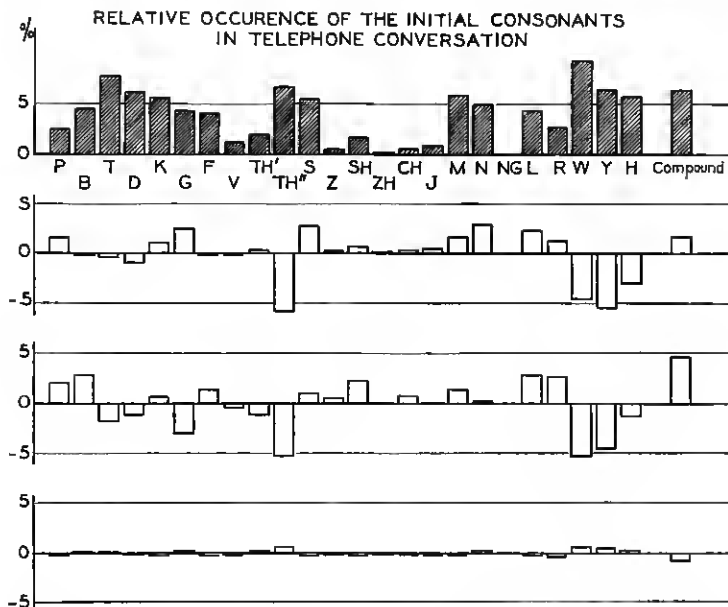


Fig. 3—The relative occurrence of initial consonants—effects of restricting the word list.

Line I—Relative occurrence of initial consonants for all words.

Line II—Differences resulting from omission of minor parts of speech (118 words).

Line III—Differences resulting from omission of the 100 commonest words.

Line IV—Differences resulting from omission of the 1,500 least common words.

hundred for initial consonants as in Table VIII. If the minor parts of speech are excluded before the analysis, which eliminates only 118 different words, but nearly half the total words, the resulting changes are shown on the second line. Notable decreases occur for "th," "w" and "y," which may largely be traced to the omission of "that," "they," "this," etc.; "will," "with," "would," etc.; and "you," respectively. These elisions enhance the relative contributions from "get," "see" and "know." When the 100 most common words are omitted there are also large changes, as shown by the third line. Since 50 of the 100 most common words are of the minor parts of speech the similarity of this line to the second is not surprising. The omission, on the other hand, of the 1,500 least common words, namely, those omitted from the vocabulary of Table III, changes the distribution by negligible amounts as shown in the fourth line. Since, then, the 737

commonest words seem to determine the relative frequency of the sounds of conversation, the writers are encouraged to believe that if this study were repeated on telephone calls of which a greater proportion were social rather than business in nature the analysis into sounds would be changed very little. Likewise the conclusion is drawn that if the study were prolonged tenfold so as to double the number of different words no material change in the relative frequency of the sounds would be found.

TABLE X

RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATIONS

Nouns, Verbs, Adjectives and Adverbs

<i>Vowels</i>		<i>Initial Consonants</i>		<i>Final Consonants</i>	
pen.....	10.63	S.....	8.34	r.....	14.64
pin.....	7.56	N.....	7.94	t.....	13.53
pane.....	7.19	T.....	7.55	n.....	9.99
pole.....	6.38	M.....	7.40	l.....	8.62
pawn.....	5.96	G.....	6.87	ng.....	5.10
peel.....	5.78	K.....	6.70	k.....	4.96
pine.....	5.40	L.....	6.65	m.....	4.50
pun.....	5.25	D.....	5.25	s.....	3.89
pot.....	4.59	W.....	4.86	d.....	3.66
pan.....	4.01	B.....	4.38	z.....	2.42
pull.....	2.49	R.....	4.11	p.....	2.00
pout.....	1.83	P.....	4.06	v.....	1.94
pair.....	1.58	F.....	3.88	f.....	1.28
par.....	1.53	SH.....	2.42	th''.....	.82
pool.....	1.39	TH'.....	2.41	ch.....	.81
purr.....	1.33	H.....	2.38	b.....	.73
pew.....	.49	J.....	1.33	g.....	.66
poise.....	.35	Y.....	1.25	sh.....	.57
		V.....	1.21	j.....	.24
	73.74	TH''.....	.97	th'.....	.07
		CH.....	.87	zh.....	.02
		Z.....	.55	h.....	—
		ZH.....	.03	w.....	—
		NG.....	—	y.....	—
			91.86		80.45
<i>Unaccented Vowels</i>		<i>Compounds</i>		<i>Compounds</i>	
differ.....	5.79	PR.....	1.69	nt.....	3.37
receive.....	5.73	ST.....	1.39	st.....	2.07
possible.....	4.82	TR.....	1.11	nd.....	1.66
about.....	3.96	PL.....	.58	nk.....	1.32
wanted.....	2.54	HW.....	.49	ld.....	1.31
people.....	1.76	KW.....	.44	rz.....	.98
notion.....	1.66	BL.....	.37	ks.....	.82
	26.26	SP.....	.30	kt.....	.73
		KL.....	.29	rd.....	.64
		GR.....	.27	ns.....	.54
		Others.....	1.21	Others.....	6.11
			8.14		19.55
			100.00		100.00
Total number...	50,161		40,107		37,493

For some purposes weighting lists based on the words of speech which carry the meaning are appropriate. This is approximated to by the figures given in Table X, in which the sounds are analyzed for nouns, verbs, adjectives and adverbs only. The outstanding changes in the weighting of initial consonants have just been commented on in con-

TABLE XI
RELATIVE OCCURRENCE OF SPEECH SOUNDS IN TELEPHONE CONVERSATIONS
Conversational Weighting

Note: The sounds of each word are weighted by the number of conversations in which the word is used, instead of by the total occurrences of the word.

<i>Vowels</i>		<i>Initial Consonants</i>		<i>Final Consonants</i>	
pin.....	11.22	W	8.26	r	13.87
pen.....	7.90	T	7.09	t	11.98
pan.....	6.40	M	6.69	n	10.92
peel.....	6.21	D	6.52	l	8.13
pine.....	5.95	K	5.90	m	5.43
pane.....	5.60	S	5.90	d	5.20
pole.....	5.18	L	5.44	z	5.13
pun.....	4.66	B	5.32	ng	4.05
pawn.....	4.64	H	5.31	v	3.72
pot.....	4.08	N	5.09	s	3.64
pool.....	3.40	TH''	5.01	k	3.41
pull.....	3.24	F	4.10	p	1.55
pout.....	1.89	G	4.00	f	1.41
par.....	1.33	R	3.53	th''	1.18
pair.....	1.31	Y	3.17	ch65
purr.....	1.11	P	3.09	b52
pew.....	.38	SH	2.09	g49
poise.....	.24	TH'	2.06	sh45
		V	1.43	j17
	74.74	J94	th'06
		CH74	zh02
		Z47	h	—
		ZH03	w	—
		NG	—	y	—
			92.18		81.98
<i>Unaccented Vowels</i>		<i>Compounds</i>		<i>Compounds</i>	
about.....	5.39	PR	1.27	nt	4.68
differ.....	5.35	ST	1.06	nd	2.09
receive.....	4.85	HW	1.03	st	1.43
possible.....	3.57	TR82	ts	1.14
notion.....	2.69	FR62	ld89
wanted.....	2.16	PL49	nk71
people.....	1.25	KW40	rz63
	25.26	BL30	ks61
		SP26	kt56
	100.00	KL26	rd51
Total Number...	54,656	Others.....	.131	Others.....	4.77
			7.82		18.02
			100.00		100.00
			39,924		40,993

nection with Fig. 3, line 2. The vowel weighting reflects the enhanced importance of the "e" in "pen" from the verbs "get," "tell," "send" and shows a considerable reduction in the vowel in "pool," largely from the loss of "you" and "to." The unstressed vowels, especially as in "about" are also diminished. Among the final consonants the largest change is a reduction in "z," which results from the elimination of "is," "was," "as," etc. Comparing the distributions of Table VIII and Table X as a whole, however, both show about the same degree of non-uniformity; the maximum and minimum weightings do not differ greatly.

One more type of analysis is given in Table XI. In this case the sounds of each word are weighted by the number of conversations in which the word occurred, instead of by the total number of times the word was used. This seems to be a somewhat radical change in method, involving, as it does, a considerable reduction in the weighting of the words at the head of the list. When the effect of eliminating the first 100 words entirely, shown in Fig. 3, line 3, is recalled, large changes might be expected. Actually the result is remarkably similar to the figures of Table VIII. The relatively diminished importance of "you" and "to" is seen in the vowel list, of "you" again among the initial consonants, and of "it," "that" and "get" in the list of final consonants. The range covered by the relative weightings is still much the same as in Tables VIII and X.

COMPARISONS WITH WRITTEN ENGLISH

Some of the differences between the vocabularies of telephone conversation and written English have been pointed out. The effects of these differences may be seen in the relative occurrence of the sounds as shown by Figures 4, 5 and 6 for vowels, initial consonants and final consonants, respectively, using the analysis based on all the words (except articles). The upper line in each case is a graphical representation of the corresponding data of Table VIII, after certain changes have been made to put them on the same basis as the tables given by Dewey for written English. In the case of the consonants the only change needed was omission of the compound consonants. In the case of the vowels it was necessary to combine some of our classifications, since Dewey made but 17 distinctions among the vowels. We believe the combinations made are those followed by Dewey himself, as ascertained from examples given by him in his text. The phonetic symbols given in Figure 4 are those used by him. The combinations made were as follows: "pen" and "wanted"; "pane" and "pair"; "pin," "possible" and "receive"; "pun" and "purr"; "about," "differ," "people"

and "notion." The comparisons are made with Table XVI of Dewey's book, which does not include "the," and from which we have subtracted the article "a."

The outstanding differences between the vowel frequencies in telephone conversation and written matter are the excess in conversation of "about," "pine," "pool" and the deficiencies in "pau," "pin"

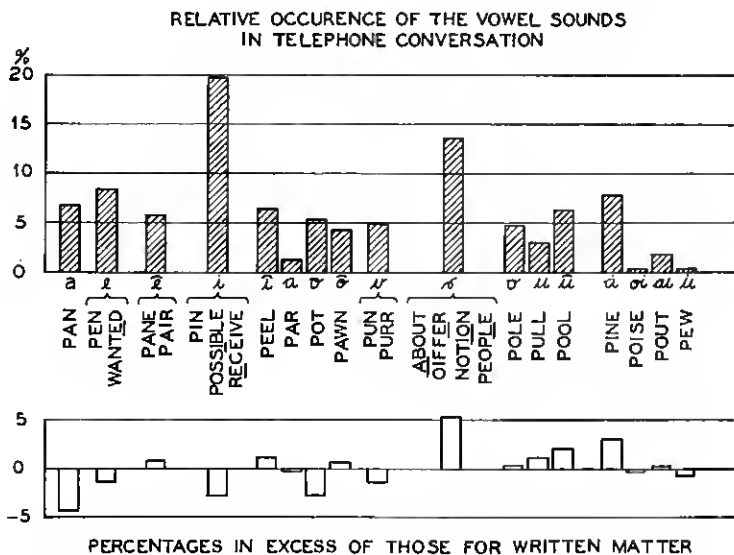


Fig. 4—Comparison with written English—relative occurrence of the vowels.

and "pot." The greater occurrences of "pine" and "pool" are almost entirely accounted for by the greater use of the words "I" and "you." The deficiencies mentioned do not, on analysis, seem to depend on one or two words, but rather on the whole vocabulary, except that of the increase in the unstressed vowel denoted by "about" nearly 1.7 per cent comes from the vowels of words which in the study of written English were classified under "pan."

Among the initial consonants (Fig. 5) the greatest change is in the occurrence of "y," which is much more frequent in conversation. This again is largely caused by the pronoun "you." Much of the increase in "g" may be traced to the greater use of "get" and "go." The sounds "w" and "t" are the most frequent sounds in written English, as well as in conversation.

Figure 6 shows that in the case of the final consonants the sounds "t" and "l" are notably more frequent in conversation than in written

matter. The increase in "t" arises almost entirely from "that," "it" and "get" which combined have a contribution about 4.9 per cent larger in conversation than in written matter. About half the increase

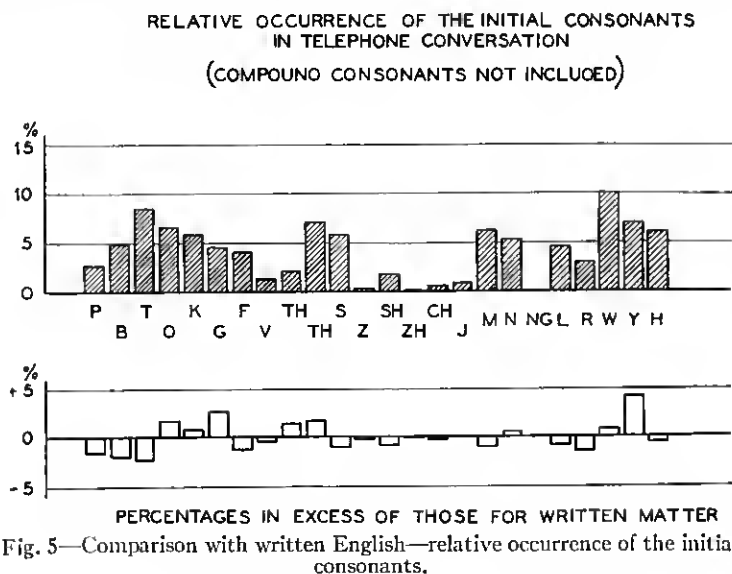


Fig. 5—Comparison with written English—relative occurrence of the initial consonants.

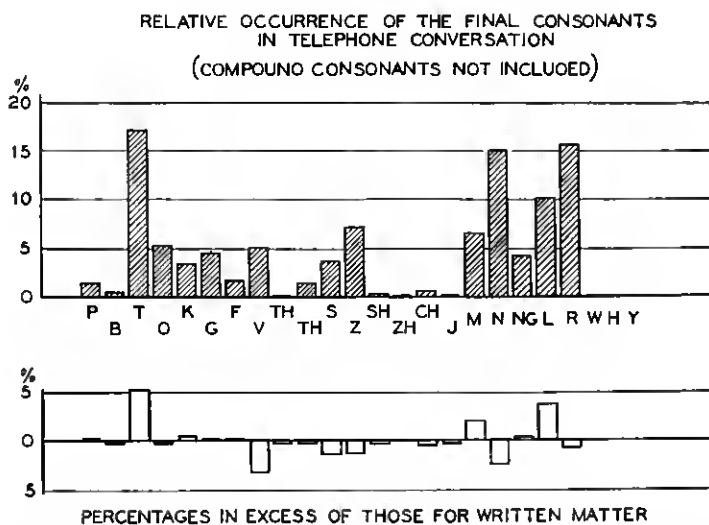


Fig. 6—Comparison with written English—relative occurrence of the final consonants. in "l" is attributable to the words "will" and "tell." Some of the deficiency in "v" may be traced to the word "of" which has a contribution 1.8 per cent greater in written matter. On the other hand the

words "have" and "give" together contribute 1.1 per cent more to conversation, so that the net difference in "v" is to be traced to small accretions from the whole vocabulary rather than a few specific words.

RELATIVE OCCURRENCE OF COMBINATIONS OF SOUNDS

A more elaborate analysis of the phonetic syllable is given in Table XII, which shows, for each vowel, the frequency of occurrence of the consonants preceding the vowel and also of the consonants which follow the vowel. The complete word list (except articles) was used as a basis. The cases in which no consonant occurs in front of the vowel are included, as well as the cases in which there is no following consonant. These figures are shown as a double column under the key word denoting the vowel sound. In each double column the figures on the left apply to initial consonants and on the right to final consonants. The figures are given in per cent, so that each column adds to 100. The consonants are grouped by phonetic classes. The table is to be read as follows: of the syllables in which the vowel sound is that in "pan," 28 per cent begin with "th" (as in "that"), 26 per cent have no initial consonant, 16 per cent begin with "h," 7 per cent with "k," 6 per cent with compound consonants, 5 per cent with "b," etc.; while 29 per cent end with compound consonants, 27 per cent with "t," etc. Where no figure is entered the occurrences were less than 0.5 per cent; where a dash is shown no combinations of the kind indicated were observed. If the figures are taken by rows instead of columns no meaning can be attached to them before they are multiplied by the relative occurrence of the different vowels.

In studying this table it is to be remembered that because the different vowels have very different frequencies of occurrence the subdivided data shown in different columns cannot be considered as equally representative. Syllables having the vowel as in "pin," for example, were present, as shown in Table VIII, to the number of $0.1027 \times 92,522$, or 9,500. The syllables in this class which begin with "t" are shown in Table XII to be 1 per cent, representing 95 occurrences. In the class having the vowel of "poise," however, there were only 176 examples, so that the 37 per cent of these syllables beginning with "p" result from only 65 occurrences.

It is to be seen that only one vowel, "pool," is preceded by a particular sound more than 50 per cent of the time, this sound naturally being "y." Six vowels are preceded by particular sounds more than 25 per cent of the time. The sounds of "pair," "purr," "par" and "differ" must be followed by "r," a blank, or a compound consonant beginning with "r," as a result of the way in which the analysis was

	PEEL	PIN	PANE	PAIR	PEN	PAN	PURR	PUN	PAR	POT	PAWN	POLE	POOL	PULL	PEW	POISE	POUT	PINE	RE- CEIVE	POSS- IBLE	WANT- ED	DIFF- ER	ABOUT	PEO- PLE	NO- TION
P	2	3	3	1	1	1	8	3	7	2	2	1	7	7	1	37	3		3	2	3	1	4	-	P
T	1	3	7	4	9	24	3	3	3	4	8	4	17	25	3	2	4		3	2	3	8	17	-	T
K	2	19	11	4	24	7	1	15	25	4	18	3	7	7	6	2	46	11	-	1	5	22	1	-	K
G	4	1	10	4	3	5	1	2	2	4	2	7	1	5	5	2	2		-	1	7	1	5	-	G
B	3	4	1	-	4	5	2	2	1	2	2	1	1	2	22	31	2		9	1	-	4	2	2	B
D	1	8	18	-	1	5	1	3	-	2	3	15	8	7	1	2	1		9	1	2	10	1	-	D
C	2	3	1	-	18	1	1	1	-	12	-	12	-	9	5	1	3		-	-	14	1	2	-	C
F	1	2	1	2	2	1	21	1	4	3	15	2	4	28	-	1	5		1	3	-	11	-	-	F
TH	1	11	-	-	-	1	1	-	-	3	2	-	-	-	-	-	-		-	2	-	-	-	-	TH
S	15	2	10	4	13	8	11	6	2	2	1	4	2	-	-	5	3	4	2	1	3	1	3	2	S
SH	3	3	2	3	3	1	-	-	-	-	-	-	-	8	-	-	-		-	7	-	-	2	-	SH
CH	3	3	2	1	2	1	-	-	2	-	-	-	-	3	-	-	-		-	-	-	-	1	2	CH
V	3	4	2	27	4	28	1	17	-	-	-	-	-	-	-	8	-		-	1	-	8	15	-	V
TH	1	3	3	44	8	28	-	7	-	-	4	-	-	-	-	-	-		-	4	5	-	-	-	TH
Z	2	5	11	-	-	1	4	-	-	-	2	8	1	3	5	7	-		1	18	23	-	-	2	Z
ZH	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	-		-	-	-	-	-	-	ZH
J	-	-	-	-	-	1	-	9	-	-	-	-	-	-	-	8	-		-	1	3	-	-	3	J
M	13	3	10	7	1	2	1	11	15	14	2	-	1	2	2	-	4		-	18	-	1	5	2	M
N	1	5	2	1	2	7	1	4	-	3	1	21	2	1	42	3	22	4	8	4	7	9	1	17	N
NG	1	3	18	2	1	15	6	9	-	45	-	8	3	3	-	1	4		-	8	2	-	4	-	NG
L	2	7	-	-	9	2	-	-	-	-	4	3	1	4	-	-	-		15	2	28	-	-	-	L
R	2	4	8	3	-	12	-	-	-	3	19	3	3	4	-	22	4	7	12	13	5	-	1	91	R
W	2	10	9	97	3	-	37	4	72	13	1	42	1	12	2	-	-		-	18	-	91	4	-	W
Y	1	-	-	-	2	-	-	-	-	-	-	-	62	6	-	-	-		-	-	-	4	-	-	Y
H	24	4	-	1	2	18	8	1	4	-	6	-	2	-	-	-	-		-	-	-	-	-	-	H
BLANK	1	39	3	2	15	26	5	28	39	46	29	15	87	5	5	20	18	58	38	56	48	47	58	68	75
COMPOUND	5	81	8	7	5	6	83	12	3	13	3	9	3	32	88	1	43	10	4	1	1	9	4	66	-
COMPOUND	4	9	4	3	20	29	29	18	23	26	26	26	3	2	2	3	3	3	4	1	32	9	1	9	52

TABLE XII

DISTRIBUTION OF CONSONANTS PRECEDING AND FOLLOWING EACH VOWEL SOUND

made, similarly for the "l" of "people" and the "n" of "notion." Aside from these no single consonant occurs as often as 50 per cent of the time after a particular vowel. With five vowels a particular consonant ends the syllable more than 25 per cent of the time. In nearly every case the most frequent combinations can be traced directly to the first 50 words of the vocabulary. Five vowels are preceded by blanks more than 50 per cent of the time and eight are followed by blanks in more than 50 per cent of the cases. The combinations of different vowels with compound consonants vary considerably in importance, ranging in the final position from practically none with the vowel of "pew" up to the vowel of "pun," which is terminated by a compound consonant 63 per cent of the time.

CONCLUSION

In concluding, a brief review is presented of the main points of interest. The paper has for its basis a list of 80,000 words obtained from telephone conversations. This list has been studied with respect to the number of different words contained in it, the relative occurrence of the different speech sounds and the combinations of sounds which form syllables. In so far as the authors know this is the first study of this type based on conversations as contrasted with written matter.

Perhaps the most striking aspect of the word list is the small number of different words contained in it, only 2,240 out of the total of 80,000. Of these 2,240 words 819 occur only once. The balance, or 1,421 words, constitute practically 99 per cent of the total words recorded; of these the 121 different words which constitute the minor parts of speech form 45,000 of the total occurrences. The pronouns "I" and "you" together occur over 7,500 times.

This intensiveness with which a small number of words is used in conversation is considerably greater than in the written English analyzed by Dewey. In conversation the 155 most frequently used words make up 80 per cent of the total occurrences; to reach the same percentage in the written English analyzed by Dewey 640 words must be included. The frequently used words of conversation are characterized, as compared with written English, by the greater prominence of certain active verbs, such as "get," "see," "know," etc., 12 of which occur in the first 50 words of conversation, while there are none in the first 50 words of written English. The most frequent words of conversation differ from written English also in the greater number of words of Latin origin which appear frequently in conversation: 11 from the first 100 of the list for conversation, as compared with two from the first 100 of written English.

The word list is characterized by a large percentage of monosyllables. Over four fifths of the 80,000 occurrences are of this type, a result largely brought about by the frequent repetition of the minor parts of speech, among which 95 per cent are monosyllables. When the words are analyzed into phonetic syllables about one fifth are found to be of the type vowel-consonant, about one fifth consonant-vowel, and a third of the type consonant-vowel-consonant.

The relative occurrences of the different speech sounds were obtained by assigning phonetic values to the sounds of the phonetic syllables and weighting each by the total number of times it was used. Twenty-five categories are used for the vowels. Seven of these are for vowels in unaccented positions, which make up, altogether, about 25 per cent of the vowels. The relative occurrences of the individual sounds differ greatly for different vowels. The range extends from about 10 per cent for the vowel of "pin," and about 8 per cent for the vowel of "pine," down to 0.3 per cent for "pew" and 0.2 per cent for "poise." Among the initial consonants 94 per cent are single sounds, and the remaining are compounds of two or more successive consonants. The range extends from about 9 per cent for "w," and about 8 per cent for "t" down to about 0.3 per cent for "z" and the slightest trace, .02 per cent for "zh." The most frequent compound initial consonant is "pr," with an occurrence of 1 per cent. Among the final consonants the compounds are somewhat more prominent, forming 16 per cent. The most frequent final consonant is "t," 14 per cent, the next is "r," 13 per cent, the range extending down to 0.1 per cent for "zh." The most frequent compound final consonant is "nt," 4.4 per cent, and the next is "nd," 2.6 per cent.

Considering the marked differences between the word lists for conversation and for written English, a comparison of the relative frequency of the speech sounds in the two cases is perhaps more remarkable for the likenesses than the differences. About the same range of percentages is covered in both cases. Certain sounds do show marked differences. Among the vowels the unaccented vowel denoted by "about" is more frequent in conversation and the vowel of "pan" less frequent. The initial "y" and the final "t" are also more frequent in conversation. Many of the differences can be traced directly to one or two words which in their frequent use are typical of conversation.

In considering the occurrence of speech sounds in telephone conversations from the point of view of their contribution to the ease or difficulty of carrying on conversations it seemed of interest to determine how the occurrence of the speech sounds was affected by changing the list in certain ways. Omission of the minor parts of speech changes the

relative occurrence of a number of the sounds materially, although the general range of percentages covered is changed very little. Omission of the 1,500 least common words has a negligible effect. When the words are weighted by the number of conversations in which they occurred, out of 500, instead of by their total occurrence, the effect on the distribution of sounds is surprisingly small, considering the radical change in method.

While the analysis into speech sounds for purposes connected with the design of telephone circuits was the real goal of this study, it is hoped that the information concerning both words and sounds will be of service also to those working in the fields of phonetics and philology.